NUWC-NPT Technical Report 11,431
28 April 2003

Combined Linear and Nonlinear Modeling of Data

Albert H. Nuttall
Surface Undersea Warfare Department

NAVSEA
NEWPORT
Undersea Warfare Center Division

Naval Undersea Warfare Center Division
Newport, Rhode Island

PREFACE

The work described in this report was funded under Project No. S0223, "Advanced Submarine Combat Systems Development Program," principal investigator Walter R. Lane (Code 2123). The sponsoring activity is the Program Executive Office Integrated Warfare Systems/ Advanced Systems and Technology Office (PEO IWS/ASTO), program manager Dr. Robert Zarnich (Code ASTO-J).

The technical reviewer for this report was David J. Pistacchio (Code 212).

Reviewed and Approved: 28 April 2003

Donald A. Aker
Head, Surface Undersea Warfare Department

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
28 April 2003

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE

Combined Linear and Nonlinear Modeling of Data

5. FUNDING NUMBERS

6. AUTHOR(S)

Albert H. Nuttall

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Undersea Warfare Center Division
1176 Howell Street
Newport, RI 02841-1708

8. PERFORMING ORGANIZATION REPORT NUMBER

TR 11,431

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Sea Systems Command
614 Sicard Street SE Stop 7017
Washington Navy Yard DC 20376-7017

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

A method is presented for reducing the dimensionality of the search space when some of the unknown parameters appear linearly in the model fit. After elimination of the linear parameters, the gradient vector and the Hessian matrix of the resultant Hermitian form are derived so that an efficient minimization procedure can be developed in multiple dimensions. A "destabilizing" term is identified in the Hessian matrix and can be dropped from the calculations if desired. This approach is expected to be more reliable; it also does not require any second-order partial derivatives, leading to fewer computations for finding the minimum in the multidimensional search space.

14. SUBJECT TERMS

Linear Modeling	Nonlinear Modeling	Combined Modeling	
Reduced Search Space	Least Squares	Gradient Vector	Hessian Matrix

15. NUMBER OF PAGES
46

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

TABLE OF CONTENTS

	Page
LIST OF ILLUSTRATIONS	ii
LIST OF ABBREVIATIONS, ACRONYMS, AND SYMBOLS	ii
INTRODUCTION	1
ERROR DEFINITION AND MINIMIZATION	5
Error Minimization	7
Gradient Vector of $\underline{E}(\theta)$	8
Hessian Matrix of $\underline{E}(\theta)$	12
SAMPLING RATES FOR PARAMETER VECTOR θ	17
QUALITY OF LEAST-SQUARES ESTIMATES	21
CHANNEL AND WAVEFORM CHARACTERIZATION	29
Time-Varying Impulse Response	29
Moving Source and Moving Receiving Elements	30
Transmitted and Received Waveforms	34
SUMMARY	39
REFERENCES	R-1

LIST OF ILLUSTRATIONS

Figure		Page
1	Conditionally-Minimum Error $\underline{E}(\theta_1)$	17

LIST OF ABBREVIATIONS, ACRONYMS, AND SYMBOLS

a	Amplitude vector $[a_1 \cdots a_K]^T$, equation (4)
$\underline{a}(\theta)$	Conditionally-optimum amplitude vector, equation (12)
a_k	k-th amplitude coefficient, equation (3)
A	Amplitude of signal
$b_k(x,\theta)$	k-th basis function, equation (3)
$B(\theta)$	Basis function matrix, N×K, equation (8)
$B_m(\theta)$	Partial derivative with respect to θ_m, equation (21)
c	Speed of propagation, equation (81)
$c_m(\theta)$	Complex component vector, equation (22)
c_{nm}	Covariance matrix elements, equation (70)
d	Data vector $[d_1 \cdots d_N]^T$, equation (1)
$D_m(t)$	Discriminant, equations (88), (90)
$e(\theta,a)$	Error vector, equation (6)
$e_n(\theta,a)$	n-th error component, equation (5)
$\underline{e}(\theta)$	Conditionally-minimum error vector, equation (25)
$e(t)$	Envelope waveform, equation (96)
$E(\theta,a)$	Total scalar error, equation (9)
$\underline{E}(\theta)$	Conditionally-minimum error, equation (14)
$\underline{E}_m(\theta)$	Partial derivative with respect to θ_m, equation (15)
$\underline{E}_{\underline{mm}}(\theta)$	Hessian matrix, equations (41) and (43)

LIST OF ABBREVIATIONS, ACRONYMS, AND SYMBOLS (Cont'd)

$f(\theta,a)$	Fitting vector, equation (7)
$f_n(\theta,a)$	n-th fitting function, equation (3)
$f(t_1)$	Reception time t_2, equation (76)
$G_E(\phi)$	Gradient vector, equation (49)
$h(t;t_1)$	Time-varying impulse response, equation (78)
H	Hermitian transpose, equation (9)
$H_E(\phi)$	Hessian matrix, equation (49)
I	In-phase component amplitude, equation (96)
K	Number of basis functions, equation (3)
M	Number of components of vector θ, equation (2)
MSE	Mean square error, equation (70)
N	Number of data values, equation (1)
N_s	Number of spatial samples
N_t	Number of temporal samples
P(T)	Position of source at time T, equation (103)
$q_m(t)$	Auxiliary time function, equation (86)
Q	Auxiliary matrix, PxP, equation (63)
Q	Quadrature component amplitude, equation (96)
r(t)	Received waveform, equation (79)
$r_m(t)$	Received waveform at m-th element, equation (97)
Re	Real part, equation (17)
s(t)	Signal waveform
t	Time
T	Transpose, equation (1)
T	Transmission time, equation (101)

LIST OF ABBREVIATIONS, ACRONYMS, AND SYMBOLS (Cont'd)

$u(t)$	Transmitted waveform, equations (79), (96)
v	Receiver velocity, equation (83)
x, x_n	General location variable, equation (3)
$x(t)$	Source location in x, equations (81), (82)
$y(t)$	Source location in y, equations (81), (82)
$\mathbf{w}_n$	Additive noise, equation (64)
$z(t)$	Source location in z, equations (81), (82)
$\cdot$	Superscript dot: derivative with respect to time
$'$	Prime: derivative, equation (79)
$\sim$	Superscript tilda: inverse function, equation (77)
boldface	Random variable or random vector
α	Auxiliary variable, equation (86)
$\boldsymbol{\alpha}_j$	Auxiliary variables, equation (56)
$\boldsymbol{\beta}_j$	Auxiliary variables, equation (58)
$\boldsymbol{\beta}(\theta)$	Complex vector, Kx1, equation (10)
$\beta_m(t)$	Auxiliary time function, equation (86)
$\boldsymbol{\beta}_m(\theta)$	Partial derivative with respect to θ_m, equation (15)
δ	Delta function, equation (76)
δ_{nm}	Kronecker delta, equation (71)
Δ_1	Sampling increment, figure 1
Δ_m	Sampling increment in θ_m
$\gamma(\theta)$	Complex matrix, KxK, equation (10)
$\gamma_m(\theta)$	Partial derivative with respect to θ_m, equation (15)

LIST OF ABBREVIATIONS, ACRONYMS, AND SYMBOLS (Cont'd)

$\gamma_m(t)$	Auxiliary time function, equation (86)
ϕ	Parameter vector $[\theta;a]$, P×1, equation (48)
σ_w	Standard deviation of noise w, equation (71)
θ	Parameter vector $[\theta_1 \cdots \theta_M]^T$, equation (2)
τ	Time delay
ω	Source radian frequency, equation (96)

COMBINED LINEAR AND NONLINEAR MODELING OF DATA

INTRODUCTION

In numerous practical applications, data are recorded for observation and scrutiny. For example, several receiving elements of an array may be observed over a time interval of interest in an effort to detect the presence of a source, estimate its location and speed, and characterize some of its attributes, such as source frequency. At the same time, there may be additional unknowns, such as the in-phase and quadrature amplitudes of multiple arrivals at the receiving elements, perhaps via direct and/or surface reflection paths.

When a model of the received signal(s) in the available measured data from the source is specified or adopted, the model generally will contain some unknown linear parameters and some unknown nonlinear parameters. For example, the in-phase and quadrature amplitudes of the received signal components will appear linearly in the source model waveform, while the source location, speed, and center frequency will appear nonlinearly in the particular source model waveform. The exact nonlinear functions depend on the configuration and geometry of the situation of interest. An example of a received signal waveform is $A\ s(t-\tau)$, where amplitude A appears linearly, while delay τ appears nonlinearly, that is, inside the function $s(\ )$.

To determine the source characteristics, it is necessary to estimate all the unknown parameters from the available data; however, this can often be accomplished via a sequential approach. For example, if the nonlinear model parameter values are initially hypothesized, an analytic solution for the conditionally-optimum linear model parameter values may be obtainable analytically through minimization of some error criterion. The conditionally-optimum linear parameters can then be substituted back into the error criterion, resulting in the reduced-dimension conditionally-minimum error, thereby often significantly reducing the size of the resultant search problem for the best (nonlinear) parameter values of the model. Reducing the dimensionality of the search space is an extremely beneficial step in terms of the amount of execution time required to find the minimum error.

When this sequential approach to a least-magnitude-squares error minimization is taken, the conditionally-minimum error takes on a Hermitian form with dimension equal to the number of nonlinear model parameters in the original error definition. To speed up the search for the global minimum in this (possibly) high-dimensional space, it is useful to be able to compute the gradient vector (slopes) and the Hessian matrix (curvatures) of the conditionally-minimum error at any point in the nonlinear search space. These quantities are indicative of the direction and step size that the next iterate for the minimum error should take.

When this task is undertaken here, the Hessian matrix is found to contain a "destabilizing" term. A similar quantity for the brute-force minimization of the original <u>total</u> squared error is presented and discussed in reference 1 (page 683, equations (15.5.7) and (15.5.11)). However, that earlier approach took no advantage of the fact that the linearly-appearing model parameters can and should be eliminated analytically, thereby significantly reducing the dimensionality and execution time of the nonlinear search for the global minimum of the specified error criterion. The current modified Hessian matrix is not as simple as that cited in reference 1; however, once again, no second-order derivative terms of the basis functions are involved, thereby reducing the amount of analytical and computer effort required.

The use of the gradient vector and the Hessian matrix of the conditionally-minimum error is of limited utility if the nonlinear search procedure does not start in the high-dimensional error valley containing the global minimum. This observation strongly suggests that a considerable fraction of the search effort should be devoted initially to locating this correct error valley, at least coarsely, before beginning a possibly wasteful time-consuming fine-grained search for a local error minimum.

This report also addresses the issue of rates of sampling in the various dimensions of the nonlinear model variables since

that issue directly affects the amount of time devoted to the initial coarse search for the proper error valley.

ERROR DEFINITION AND MINIMIZATION

The measured data set is presumed to contain N values:

$$d = [d_1 \cdots d_N]^T , \tag{1}$$

where $\{d_n\}$ could be complex. These N (random) data values could consist, for example, of N_s spatial samples (array elements) and N_t time samples, in which case, $N = N_s N_t$.

The basis set of functions consists of K known complex functions $\{b_k(x,\theta)\}$ for k=1:K, where x is a general real location variable (space and time) at which the data are measured; that is, data value d_n is measured at known location x_n, n=1:N. The M×1 nonlinear unknown complex parameter vector is

$$\theta = [\theta_1 \cdots \theta_M]^T . \tag{2}$$

These basis functions are weighted and summed to form a fit to the measured data according to

$$f_n(\theta,a) = \sum_{k=1}^{K} a_k \, b_k(x_n,\theta) \quad \text{for } n=1:N , \tag{3}$$

where the K×1 unknown complex amplitude vector is

$$a = [a_1 \cdots a_K]^T . \tag{4}$$

The fitting function, equation (3), is linear in the K×1 amplitude vector a, but it is nonlinear in the M×1 parameter vector θ.

An instantaneous error between fit and data is defined as

$$e_n(\theta,a) = f_n(\theta,a) - \mathbf{d}_n \quad \text{for } n=1:N \,, \tag{5}$$

or, in vector notation,

$$\mathbf{e}(\theta,a) = [\mathbf{e}_1(\theta,a) \cdots \mathbf{e}_N(\theta,a)]^T = \mathbf{f}(\theta,a) - \mathbf{d} \,. \tag{6}$$

By use of equation (3), the Nx1 fitting vector f(θ,a) can be expressed as

$$\mathbf{f}(\theta,a) = [f_1(\theta,a) \cdots f_N(\theta,a)]^T = B(\theta)\ a \,, \tag{7}$$

where the NxK basis function matrix B(θ) is given by

$$B(\theta) = \begin{bmatrix} b_1(x_1,\theta) & \cdots & b_K(x_1,\theta) \\ \vdots & & \vdots \\ b_1(x_N,\theta) & \cdots & b_K(x_N,\theta) \end{bmatrix} . \tag{8}$$

Observe that the (θ,a) dependency of f(θ,a) in equation (7) has factored into the product of a matrix dependent only on θ times the amplitude vector a.

The total scalar error of the complete fit is defined as

$$\begin{aligned} E(\theta,a) &= \mathbf{e}(\theta,a)^H\ \mathbf{e}(\theta,a) = [B(\theta)\ a - \mathbf{d}]^H\ [B(\theta)\ a - \mathbf{d}] \\ &= a^H\ \gamma(\theta)\ a - a^H\ \boldsymbol{\beta}(\theta) - \boldsymbol{\beta}(\theta)^H\ a + \mathbf{d}^H\ \mathbf{d} \,, \end{aligned} \tag{9}$$

where complex matrices

$$\gamma(\theta) = B(\theta)^H\ B(\theta) = \gamma(\theta)^H \,, \qquad \boldsymbol{\beta}(\theta) = B(\theta)^H\ \mathbf{d} \,. \tag{10}$$

Matrix $\gamma(\theta)$ is KxK and is positive definite, while vector $\boldsymbol{\beta}(\theta)$ is Kx1. These are relatively small matrices, at least compared to the dimensionality N of the measured data $\{\mathbf{d}_n\}$.

ERROR MINIMIZATION

The scalar error in equation (9) can be expressed as

$$\mathbf{E}(\theta,a) = [a - \gamma(\theta)^{-1}\ \boldsymbol{\beta}(\theta)]^H\ \gamma(\theta)\ [a - \gamma(\theta)^{-1}\ \boldsymbol{\beta}(\theta)]$$
$$+\ \mathbf{d}^H\ \mathbf{d} - \boldsymbol{\beta}(\theta)^H\ \gamma(\theta)^{-1}\ \boldsymbol{\beta}(\theta)\ . \tag{11}$$

Since matrix $\gamma(\theta)$ is positive definite for all θ, the best (random) amplitude vector a to minimize $\mathbf{E}(\theta,a)$ is given by

$$\underline{a}(\theta) = \gamma(\theta)^{-1}\ \boldsymbol{\beta}(\theta)\ , \tag{12}$$

which depends on the particular hypothesized vector value of θ. As nonlinear parameter vector θ changes, this conditionally-optimum amplitude vector $\underline{a}(\theta)$ also varies.

An equivalent form to equation (12) is

$$\gamma(\theta)\ \underline{a}(\theta) = \boldsymbol{\beta}(\theta) \quad \text{or} \quad B(\theta)^H\ B(\theta)\ \underline{a}(\theta) = B(\theta)^H\ \mathbf{d}\ , \tag{13}$$

which is recognized as the normal form of the equations that result from the least-squares procedure for the fit $B(\theta)\ a \sim \mathbf{d}$. This result is also evident from the upper line of equation (9). In MATLAB notation, solution $\underline{a}(\theta) = B(\theta)\backslash\mathbf{d}$ instead of equation (12).

Upon substitution of solution $\underline{a}(\theta)$ into equation (11), the conditionally-minimum scalar error becomes

$$\underline{E}(\theta) \equiv E(\theta,\underline{a}(\theta)) = \mathbf{d}^H \mathbf{d} - \boldsymbol{\beta}(\theta)^H \gamma(\theta)^{-1} \boldsymbol{\beta}(\theta)$$

$$= \mathbf{d}^H \mathbf{d} - \boldsymbol{\beta}(\theta)^H \underline{a}(\theta) . \tag{14}$$

This quantity $\underline{E}(\theta)$ is the minimum total scalar error for a given or hypothesized parameter vector value θ. Vector θ is presumed real henceforth. $\underline{E}(\theta)$ must now be further minimized by choice of θ. This must be accomplished by a search in the M-dimensional space of θ. A brute-force search directly on Hermitian form $\boldsymbol{\beta}(\theta)^H \gamma(\theta)^{-1} \boldsymbol{\beta}(\theta)$, for its maximum in θ, is one possible alternative. However, the M-dimensional search is often accomplished more quickly by using the gradient vector and the Hessian matrix of the conditionally-minimum error $\underline{E}(\theta)$, at least when the correct error valley of $\underline{E}(\theta)$ has already been located.

GRADIENT VECTOR OF $\underline{E}(\theta)$

A partial derivative with respect to the m-th component θ_m of vector θ in equation (2) will be denoted by a subscript m, m=1:M. Then, there follows from equation (14), scalars

$$\underline{E}_m(\theta) = \frac{\partial}{\partial \theta_m} \underline{E}(\theta) = - \boldsymbol{\beta}_m(\theta)^H \gamma(\theta)^{-1} \boldsymbol{\beta}(\theta) - \boldsymbol{\beta}(\theta)^H \gamma(\theta)^{-1} \boldsymbol{\beta}_m(\theta)$$

$$+ \boldsymbol{\beta}(\theta)^H \gamma(\theta)^{-1} \gamma_m(\theta) \gamma(\theta)^{-1} \boldsymbol{\beta}(\theta) \quad \text{for } m=1:M . \tag{15}$$

On the other hand, consider from equation (9) the quantity

$$\frac{\partial}{\partial \theta_m} E(\theta,a) = a^H \gamma_m(\theta)\ a - a^H \boldsymbol{\beta}_m(\theta) - \boldsymbol{\beta}_m(\theta)^H a \ , \tag{16}$$

where amplitude vector a still has a general (but fixed) value. Upon now setting general value a equal to the conditionally-optimum value $\underline{a}(\theta)$, equation (16) yields

$$\frac{\partial}{\partial \theta_m} E(\theta,a)\bigg|_{a=\underline{a}(\theta)} = \underline{a}(\theta)^H \gamma_m(\theta)\ \underline{a}(\theta) - 2\ \mathrm{Re}\left(\underline{a}(\theta)^H \boldsymbol{\beta}_m(\theta)\right) \tag{17}$$

$$= \boldsymbol{\beta}(\theta)^H \gamma(\theta)^{-1} \gamma_m(\theta)\ \gamma(\theta)^{-1} \boldsymbol{\beta}(\theta) - 2\ \mathrm{Re}\left(\boldsymbol{\beta}(\theta)^H \gamma(\theta)^{-1} \boldsymbol{\beta}_m(\theta)\right). \tag{18}$$

This result is identical to equation (15); that is, for all θ, the gradient components of $\underline{E}(\theta)$ satisfy

$$\underline{E}_m(\theta) = \frac{\partial}{\partial \theta_m} E(\theta,\underline{a}(\theta)) = \frac{\partial}{\partial \theta_m} E(\theta,a)\bigg|_{a=\underline{a}(\theta)} \quad \text{for } m=1:M \ , \tag{19}$$

where $\underline{a}(\theta) = \gamma(\theta)^{-1} \boldsymbol{\beta}(\theta)$, and the latter quantities are given by equation (10). Thus, the setting of general a to the optimum $\underline{a}(\theta)$ can be done either before or after taking the partial derivative of the total error $E(\theta,a)$ with respect to θ_m.

The forms in equations (15) and (18) are not immediately useful because $\gamma_m(\theta)$ and $\boldsymbol{\beta}_m(\theta)$ have not yet been evaluated. From equations (17) and (19),

$$\underline{E}_m(\theta) = \underline{a}^H \gamma_m\ \underline{a} - 2\ \mathrm{Re}\left(\underline{a}^H \boldsymbol{\beta}_m\right) \quad \text{for } m=1:M \ , \tag{20}$$

where θ has been temporarily suppressed from the right side of this equation. From equation (10), however,

$$\gamma_m = B_m^H B + B^H B_m \, , \qquad \boldsymbol{\beta}_m = B_m^H \mathbf{d} \quad \text{for } m=1:M \, . \tag{21}$$

At this point, define the N×1 vectors

$$\mathbf{c}_m = B_m \, \underline{\mathbf{a}} \quad \text{for } m=1:M \, . \tag{22}$$

Then, equation (20) yields

$$\underline{\mathbf{E}}_m(\theta) = \underline{\mathbf{a}}^H \, (B_m^H B + B^H B_m) \, \underline{\mathbf{a}} - \underline{\mathbf{a}}^H B_m^H \mathbf{d} - \mathbf{d}^H B_m \, \underline{\mathbf{a}}$$

$$= \mathbf{c}_m^H B \, \underline{\mathbf{a}} + \underline{\mathbf{a}}^H B^H \mathbf{c}_m - \mathbf{c}_m^H \mathbf{d} - \mathbf{d}^H \mathbf{c}_m \, . \tag{23}$$

Upon restoration of the θ dependence, this becomes, for m=1:M,

$$\underline{\mathbf{E}}_m(\theta) = \frac{\partial}{\partial \theta_m} E(\theta, \underline{\mathbf{a}}(\theta)) = 2 \, \mathrm{Re}\Big(\mathbf{c}_m(\theta)^H \, [B(\theta) \, \underline{\mathbf{a}}(\theta) - \mathbf{d}]\Big) \, . \tag{24}$$

An alternative form is available from equations (6) and (7) in terms of the conditionally-minimum error vector:

$$\underline{\mathbf{e}}(\theta) \equiv \mathbf{e}(\theta, \underline{\mathbf{a}}(\theta)) = B(\theta) \, \underline{\mathbf{a}}(\theta) - \mathbf{d} \, , \tag{25}$$

namely,

$$\underline{\mathbf{E}}_m(\theta) = 2 \, \mathrm{Re}\Big(\mathbf{c}_m(\theta)^H \, \underline{\mathbf{e}}(\theta)\Big) \quad \text{for } m=1:M \, . \tag{26}$$

This form may be useful computationally, in that the N×1 error vector $\underline{\mathbf{e}}(\theta)$ is independent of m and needs to be calculated only once at each θ of interest; on the other hand, N×1 vector $\mathbf{c}_m(\theta)$ depends on both m and θ.

A quantity that plays a prominent role in the calculation of the gradient of $\underline{E}(\theta)$ is the N×K partial derivative matrix $B_m(\theta)$ that arose in equation (21). From equation (8), it follows that

$$B_m(\theta) = \frac{\partial}{\partial\theta_m}B(\theta) = \begin{bmatrix} \frac{\partial}{\partial\theta_m} b_1(x_1,\theta) & \cdots & \frac{\partial}{\partial\theta_m} b_K(x_1,\theta) \\ \vdots & & \vdots \\ \frac{\partial}{\partial\theta_m} b_1(x_N,\theta) & \cdots & \frac{\partial}{\partial\theta_m} b_K(x_N,\theta) \end{bmatrix} \quad \text{for } m=1:M. \tag{27}$$

The basic derivatives required are $\partial b_k(x,\theta)/\partial\theta_m$ for k=1:K, m=1:M, which must be evaluated at each N×1 location vector x and M×1 parameter vector θ of interest. Then, from equation (22), the quantities

$$c_m(\theta) = B_m(\theta)\ \underline{a}(\theta) \quad \text{for } m=1:M \tag{28}$$

can be evaluated. Finally, their use in equation (24) allows for calculation of the gradient vector of $\underline{E}(\theta)$. Alternatively, the combination of equations (26) and (28) yields

$$\underline{E}_m(\theta) = 2\ \mathrm{Re}\left(\underline{a}(\theta)^H\ B_m(\theta)^H\ \underline{e}(\theta)\right) \quad \text{for } m=1:M\ . \tag{29}$$

The equality of first-order partial derivatives in equation (19) does not extend to second order; that is,

$$\frac{\partial^2}{\partial\theta_m\ \partial\theta_{\underline{m}}} E(\theta,a)\Bigg|_{a=\underline{a}(\theta)} \neq \frac{\partial^2}{\partial\theta_m\ \partial\theta_{\underline{m}}} E(\theta,\underline{a}(\theta))\ . \tag{30}$$

For example, with N = 1, K = 1, M = 1, it follows that

$$e(\theta_1,a_1) = a_1\ b(x_1,\theta_1) - d_1 \equiv a_1\ b - d_1 \ ,$$

$$E(\theta_1,a_1) = |a_1\ b - d_1|^2 \ , \quad \underline{a}_1 = d_1/b \ ,$$

$$\underline{E}(\theta_1) = E(\theta_1,\underline{a}_1) = 0 \quad \text{for all } \theta_1 \ . \tag{31}$$

On the other hand,

$$\frac{\partial}{\partial\theta_1} E(\theta_1,a_1) = a_1\ (a_1\ b - d_1)^*\ b_{\theta_1} + a_1^*\ (a_1\ \bar{b} - d_1)\ b_{\theta_1}^* ,$$

$$\frac{\partial^2}{\partial\theta_1^2} E(\theta_1,a_1) = 2\ |a_1|^2\ |b_{\theta_1}|^2 + 2\ \mathrm{Re}\left(a_1\ (a_1\ b - d_1)^*\ b_{\theta_1\theta_1}\right) ,$$

$$\left.\frac{\partial^2}{\partial\theta_1^2} E(\theta_1,a_1)\right|_{a_1=\underline{a}_1(\theta_1)} = 2\ |\underline{a}_1^2\ b_{\theta_1}^2| = 2\ |d_1\ b_{\theta_1}/b|^2 > 0. \tag{32}$$

This positive value is in contrast to all the zero derivatives that will result from equation (31) for all values of θ_1. Thus, there is no shortcut at the second-order level corresponding to that in equation (19) at the first-order level. To determine the Hessian matrix of $\underline{E}(\theta)$, it is necessary to deal directly with forms (17), (24), (26), or (29).

HESSIAN MATRIX OF $\underline{E}(\theta)$

Consider the second-order partial derivative of the original scalar error $E(\theta,a)$ with respect to the components of θ; namely, from equation (16),

$$\frac{\partial^2}{\partial\theta_m\ \partial\theta_{\underline{m}}} E(\theta,a) = a^H\ \gamma_{m\underline{m}}(\theta)\ a - 2\ \mathrm{Re}\left(a^H\ \boldsymbol{\beta}_{m\underline{m}}(\theta)\right) \quad \text{for } m,\underline{m}=1:M. \quad (33)$$

Substitution of the optimum amplitude vector $\underline{a}(\theta) = \gamma(\theta)^{-1}\ \boldsymbol{\beta}(\theta)$ at this stage then yields

$$\left.\frac{\partial^2}{\partial\theta_m\ \partial\theta_{\underline{m}}} E(\theta,a)\right|_{a=\underline{a}(\theta)} = \underline{a}(\theta)^H\ \gamma_{m\underline{m}}(\theta)\ \underline{a}(\theta) - 2\ \mathrm{Re}\left(\underline{a}(\theta)^H\ \boldsymbol{\beta}_{m\underline{m}}(\theta)\right) =$$

$$\boldsymbol{\beta}(\theta)^H\ \gamma(\theta)^{-1}\ \gamma_{m\underline{m}}(\theta)\ \gamma(\theta)^{-1}\ \boldsymbol{\beta}(\theta) - 2\ \mathrm{Re}\left(\boldsymbol{\beta}(\theta)^H\ \gamma(\theta)^{-1}\ \boldsymbol{\beta}_{m\underline{m}}(\theta)\right) \quad (34)$$

for $m,\underline{m}=1:M$. However, this quantity is <u>not</u> the $m,\underline{m}$ term of the Hessian matrix of conditionally-minimum error $\underline{E}(\theta)$ in equation (14), as will be seen now.

From equations (17) and (19), suppressing θ on the right side temporarily,

$$\underline{E}_m(\theta) = \frac{\partial}{\partial\theta_m} E(\theta,\underline{a}(\theta)) = \underline{a}^H\ \gamma_m\ \underline{a} - \underline{a}^H\ \boldsymbol{\beta}_m - \boldsymbol{\beta}_m^H\ \underline{a} \quad \text{for } m=1:M\ . \quad (35)$$

Then, the $m,\underline{m}$ term of the Hessian matrix of $\underline{E}(\theta) = E(\theta,\underline{a}(\theta))$ is

$$\underline{E}_{m\underline{m}}(\theta) \equiv \frac{\partial^2}{\partial\theta_m\ \partial\theta_{\underline{m}}} \underline{E}(\theta) = \frac{\partial^2}{\partial\theta_m\ \partial\theta_{\underline{m}}} E(\theta,\underline{a}(\theta)) = \underline{a}^H\ \gamma_{m\underline{m}}\ \underline{a}$$

$$+\ \underline{a}_{\underline{m}}^H\ \gamma_m\ \underline{a} + \underline{a}^H\ \gamma_m\ \underline{a}_{\underline{m}} - \underline{a}_{\underline{m}}^H\ \boldsymbol{\beta}_m - \boldsymbol{\beta}_m^H\ \underline{a}_{\underline{m}} - \underline{a}^H\ \boldsymbol{\beta}_{m\underline{m}} - \boldsymbol{\beta}_{m\underline{m}}^H\ \underline{a}\ . \quad (36)$$

From equation (13), however,

$$\gamma\ \underline{a} = \boldsymbol{\beta}\ , \qquad \gamma_m\ \underline{a} + \gamma\ \underline{a}_m = \boldsymbol{\beta}_m\ , \quad (37)$$

and it follows that

$$\gamma_m \, \underline{a} = \beta_m - \gamma \, \underline{a}_m \quad \text{for } m=1:M . \tag{38}$$

Substitution in equation (36) and simplification results in

$$E_{m\underline{m}}(\theta) = \underline{a}^H \, \gamma_{m\underline{m}} \, \underline{a} - 2 \, \mathrm{Re}\left(\underline{a}^H \, \beta_{m\underline{m}}\right) - 2 \, \mathrm{Re}\left(\underline{a}_m^H \, \gamma \, \underline{a}_{\underline{m}}\right) \tag{39}$$

for $m,\underline{m}=1:M$. There is an additional (last) term in equation (39) that is absent in equation (34). Thus, the attempted shortcut in equations (33) and (34) is incorrect.

To determine the correct Hessian matrix, equation (10) is used to obtain the relations

$$\beta_m = B_m^H \, d \, , \quad \beta_{m\underline{m}} = B_{m\underline{m}}^H \, d \, , \quad \gamma_m = B_m^H \, B + B^H \, B_m \, ,$$

$$\gamma_{m\underline{m}} = B_{m\underline{m}}^H \, B + B_m^H \, B_{\underline{m}} + B_{\underline{m}}^H \, B_m + B^H \, B_{m\underline{m}} \, . \tag{40}$$

Substitution in equation (39) and simplification yield the exact result for the terms of the Hessian matrix of $\underline{E}(\theta)$ as

$$E_{m\underline{m}}(\theta) = 2 \, \mathrm{Re}\left((B_m \, \underline{a})^H \, (B_{\underline{m}} \, \underline{a})\right) - 2 \, \mathrm{Re}\left((B \, \underline{a}_m)^H \, (B \, \underline{a}_{\underline{m}})\right)$$

$$+ \, 2 \, \mathrm{Re}\left(\underline{a}^H \, B_{m\underline{m}}^H \, (B \, \underline{a} - d)\right) \quad \text{for } m,\underline{m}=1:M . \tag{41}$$

From equation (13), it follows that

$$B^H \, (B \, \underline{a} - d) = 0 \ (K\times 1 \text{ vector}) . \tag{42}$$

Although this $K\times 1$ relation does not require that $B \, \underline{a} - d = 0$ ($N\times 1$ vector), it suggests that $N\times 1$ vector $B \, \underline{a} - d$ will be small; in fact, from equation (25), this $N\times 1$ vector is just the

conditionally-minimum error vector $\underline{e}(\theta)$ at the θ value of interest. The N component terms of $\underline{e}(\theta)$ can have either polarity and will be unrelated to the second-order partial derivatives $\{B_{m\underline{m}}\}$, which are solely model dependent. Therefore, the last scalar, that is, the bottom line of equation (41), will tend to average out to zero and could be dropped if desired. This destabilizing term is very similar to that in reference 1 (page 683) with relation to their simpler least-squares problem in equations (15.5.5) and (15.5.11). Therefore, a reasonable approximation to the Hessian matrix of $\underline{E}(\theta)$ is afforded by

$$\underline{E}^{(a)}_{m\underline{m}}(\theta) \equiv 2\ \mathrm{Re}\left((B_m\ \underline{a})^H\ (B_{\underline{m}}\ \underline{a})\right) - 2\ \mathrm{Re}\left((B\ \underline{a}_m)^H\ (B\ \underline{a}_{\underline{m}})\right) \qquad (43)$$

for $m,\underline{m}=1:M$. This result applies for all θ. No second-order partial derivatives of the basis functions $\{b_k(x,\theta)\}$ are required to evaluate this approximate Hessian matrix (43). Only the first-order partial derivatives indicated in equation (27) need be evaluated.

Before the exact result (41) or approximation (43) can be used, the quantities $\{\underline{a}_m\}$ for $m=1:M$ need to be determined. From equation (37), it follows that

$$\gamma\ \underline{a}_m = \boldsymbol{\beta}_m - \gamma_m\ \underline{a} \quad \text{for } m=1:M \quad (K\times1 \text{ vectors})\ . \qquad (44)$$

Use of this relation in equation (43) yields an alternative form for the elements of the approximate Hessian matrix as

$$\underline{E}_{\underline{m}m}^{(a)}(\theta) = 2\ \mathrm{Re}\left((B_m\ \underline{a})^H\ (B_{\underline{m}}\ \underline{a})\right)$$

$$- 2\ \mathrm{Re}\left((\boldsymbol{\beta}_m - \gamma_m\ \underline{a})^H\ \gamma^{-1}\ (\boldsymbol{\beta}_{\underline{m}} - \gamma_{\underline{m}}\ \underline{a})\right) . \qquad (45)$$

To summarize, the N×1 data vector **d** is given by equation (1) while the N×K basis function matrices B and B_m are given by equations (8) and (27), respectively. The γ and $\boldsymbol{\beta}$ matrices are given by equation (10), while the conditionally-optimum amplitude vector $\underline{a}$ is presented in equation (12). The partial derivatives of $\boldsymbol{\beta}$ and γ are

$$\boldsymbol{\beta}_m(\theta) = B_m(\theta)^H\ \mathbf{d} \quad \text{for } m=1:M\ , \qquad (46)$$

and

$$\gamma_m(\theta) = B_m(\theta)^H\ B(\theta) + B(\theta)^H\ B_m(\theta) = \gamma_m(\theta)^H \quad \text{for } m=1:M\ . \qquad (47)$$

Each $\boldsymbol{\beta}_m(\theta)$ is a K×1 vector, while each $\gamma_m(\theta)$ is a K×K matrix. Here, K is the number of basis functions $\{b_k(x,\theta)\}$, M is the size of nonlinear real parameter vector θ, and N is the total number of data points $\{\mathbf{d}_n\}$ to be fit.

SAMPLING RATES FOR PARAMETER VECTOR Θ

The derivations in the previous section for the gradient vector and the Hessian matrix of the conditionally-minimum error $\underline{E}(\Theta)$ are useful only when the Θ valley of $\underline{E}(\Theta)$ containing the global minimum has been located. Otherwise, a fine-grained search in any other Θ valley yields only a local minimum of $\underline{E}(\Theta)$, which could have a significantly larger value than the global minimum of $\underline{E}(\Theta)$. Therefore, a very significant fraction of the total search effort for the global minimum of $\underline{E}(\Theta)$ must be devoted to locating the correct valley in Θ, in the first place.

Consider, first, the case of $M = 1$, that is, real parameter vector Θ has just one component Θ_1. A possible sample of conditionally-minimum error $\underline{E}(\Theta_1)$ is depicted in figure 1. Parameter Θ_1 is limited to observation interval (Θ_a,Θ_b). If value Θ_c is selected as the starting point for a fine-grained search in Θ_1, the local minimum at Θ_d will be reached. On the other hand, if value Θ_e is selected as the initial search point, the global minimum in interval (Θ_a,Θ_b) will be realized at Θ_f.

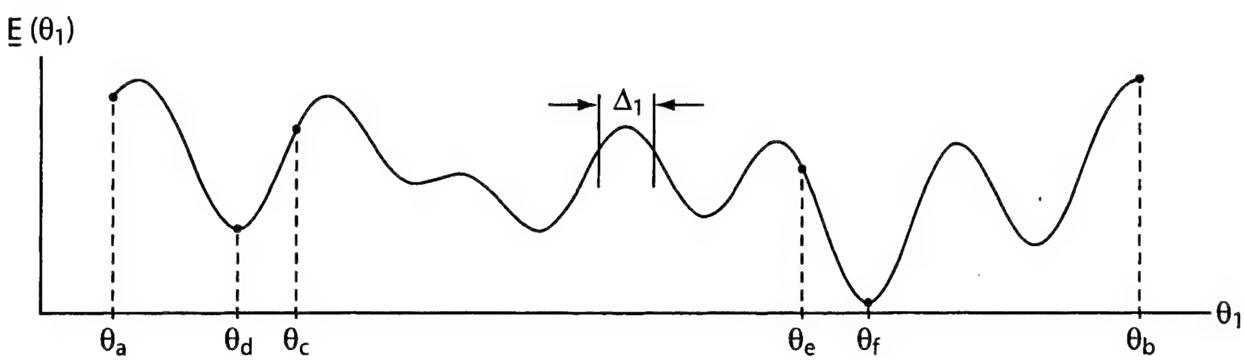

Figure 1. Conditionally-Minimum Error $\underline{E}(\Theta_1)$

To guarantee that the correct valley of $\underline{E}(\theta_1)$ is not missed during the initial coarse search in θ_1, it is necessary to sample in parameter value θ_1 with an increment of the order of Δ_1, as indicated in figure 1. That is, the Nyquist rate of variation of $\underline{E}(\theta_1)$ with θ_1 must be determined so that a sufficiently fine increment Δ_1 can be determined. Too fine a choice for Δ_1 will result in highly-dependent function value samples for $\underline{E}(\theta_1)$ and a wasteful excessive search time. Too coarse a choice for Δ_1 can cause the proper valley of $\underline{E}(\theta)$ to be missed entirely. Perhaps the safest approach is to compute a handful of samples of $\underline{E}(\theta_1)$ with some initial guess for Δ_1 and to plot the results. Cases of too fine sampling or too coarse sampling will be obvious from this plot, and a correction of Δ_1 in the proper direction can then be made. The object is to sample as coarsely in θ_1 as possible, without missing any of the valleys of $\underline{E}(\theta_1)$.

For values of M larger than 1, θ is a vector of M real components, and the search, initial as well as final, must take place in M dimensions. Perhaps the safest approach now, during the initial coarse search phase, is to hold all M components of vector θ fixed at some nominal values except for one component, say θ_m. Then, plot $\underline{E}(\theta)$ versus θ_m for a handful of samples in θ_m using some initial guess for corresponding increment Δ_m and thereby ascertain a proper value for this particular increment Δ_m. Repeat this procedure sequentially for each dimension, m=1:M, until a complete set of acceptable increments $\{\Delta_m\}$ has

been determined. Finally, conduct the full-scale coarse search in M-dimensional θ using these increments $\{\Delta_m\}$ and thereby locate the valley containing the global minimum. Whether this complete search can be accomplished in a reasonable amount of time depends on the size of M (the curse of dimensionality) and the extent of search (initial uncertainty) required on each component θ_m, m=1:M, of vector θ.

The nominal values at which M-1 of the components of θ are held fixed should hopefully be in a fairly reasonable neighborhood of the (unknown) true global minimum of $\underline{E}(\theta)$. Otherwise, the selected values of increments $\{\Delta_m\}$ could be misleading, some being too large and/or some being too small. Each initial individual one-dimensional plot of $\underline{E}(\theta)$ versus θ_m will yield some information as to whether the corresponding increment Δ_m is as valid at one end of the plot as it is at the other end of the plot. Also, since the dimensions of the parameter vector components $\{\theta_m\}$ are generally different (for example, seconds, meters, Hertz), the proper individual increments $\{\Delta_m\}$ could be very different in magnitude. Thus, this initial coarse search is extremely worthwhile and probably mandatory for efficient localization of the global minimum.

QUALITY OF LEAST-SQUARES ESTIMATES

Recall from equation (1) that **d** is the Nx1 data vector of received data, counting both space and time samples. The Nx1 fitting vector $f(\theta,a)$ in equation (7) is replaced here by $f(\phi)$, where the new parameter vector $\phi = [\theta;a]$ incorporates all the old parameter variables and is of size Px1. In this section, it is presumed that the data vector **d**, basis function matrix $B(\theta)$, amplitude vector a, and parameter vector θ are all real.

The instantaneous error vector is now $e(\phi) = f(\phi) - \mathbf{d}$ and is Nx1. The total squared error is

$$E(\phi) = e(\phi)^T e(\phi) = \sum_{n=1}^{N} e_n^2(\phi) = [\mathbf{d}^T - f(\phi)^T][\mathbf{d} - f(\phi)] \ . \qquad (48)$$

Let $\underline{\phi}$ be a local minimum of $E(\phi)$ and define Px1 difference vector $\Delta = \phi - \underline{\phi}$. By holding $\underline{\phi}$ fixed, the total error can be expanded according to

$$E(\phi) \cong E(\underline{\phi}) + G_E(\underline{\phi})^T \Delta + \frac{1}{2} \Delta^T H_E(\underline{\phi}) \Delta \quad \text{for } \phi \text{ near } \underline{\phi} \ . \qquad (49)$$

The Px1 gradient vector $G_E(\phi)$ is zero at $\phi = \underline{\phi}$. $H_E(\phi)$ is the PxP Hessian matrix of scalar error $E(\phi)$; the value of H_E required in equation (49) can be calculated once $\underline{\phi}$ has been determined.

Also, the n-th component of fitting vector $f(\phi)$ is given by

$$f_n(\phi) \cong f_n(\underline{\phi}) + G_n(\underline{\phi})^T \Delta + \frac{1}{2} \Delta^T H_n(\underline{\phi}) \Delta \quad \text{for } \phi \text{ near } \underline{\phi} \qquad (50)$$

and for n=1:N, where

$$G_n(\phi) = \left[\frac{\partial f_n(\phi)}{\partial \phi_1} \cdots \frac{\partial f_n(\phi)}{\partial \phi_P}\right]^T \tag{51}$$

is the Px1 gradient vector of scalar function $f_n(\phi)$, and

$$H_n(\phi) = \left[\frac{\partial^2 f_n(\phi)}{\partial \phi_p \ \partial \phi_q}\right] \ , \ p,q=1:P \tag{52}$$

is the PxP Hessian matrix of scalar function $f_n(\phi)$. From equation (50) follows the expansion:

$$f(\phi) = \begin{bmatrix} f_1(\phi) \\ \vdots \\ f_N(\phi) \end{bmatrix} \cong \begin{bmatrix} f_1(\underline{\phi}) \\ \vdots \\ f_N(\underline{\phi}) \end{bmatrix} + \begin{bmatrix} G_1(\underline{\phi})^T \ \Delta \\ \vdots \\ G_N(\underline{\phi})^T \ \Delta \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \Delta^T \ H_1(\underline{\phi}) \ \Delta \\ \vdots \\ \Delta^T \ H_N(\underline{\phi}) \ \Delta \end{bmatrix} . \tag{53}$$

This expansion will now be employed in equation (48), namely,

$$E(\phi) = d^T d - 2 \ f(\phi)^T d + f(\phi)^T f(\phi) \ . \tag{54}$$

The first term of interest is

$$f(\phi)^T d = \sum_{n=1}^{N} d_n \ f_n(\phi) = \sum_{n=1}^{N} d_n \ f_n(\underline{\phi}) + \sum_{n=1}^{N} d_n \ G_n(\underline{\phi})^T \Delta$$

$$+ \frac{1}{2} \sum_{n=1}^{N} d_n \ \Delta^T H_n(\underline{\phi}) \ \Delta = \alpha_0 + \alpha_1^T \Delta + \frac{1}{2} \Delta^T \alpha_2 \Delta \ , \tag{55}$$

where random variables

$$\alpha_0 = \sum_{n=1}^{N} d_n \ f_n(\underline{\phi}) \ , \quad \alpha_1 = \sum_{n=1}^{N} d_n \ G_n(\underline{\phi}) \ , \quad \alpha_2 = \sum_{n=1}^{N} d_n \ H_n(\underline{\phi}) . \tag{56}$$

The second term of interest is

$$f(\phi)^T f(\phi) = \sum_{n=1}^{N} f_n^2(\phi) \cong \sum_{n=1}^{N} \left(f_n(\underline{\phi}) + G_n(\underline{\phi})^T \Delta + \frac{1}{2} \Delta^T H_n(\underline{\phi}) \Delta\right)^2$$

$$\cong \beta_0 + 2 \beta_1^T \Delta + \Delta^T \beta_2 \Delta , \tag{57}$$

where the quantities

$$\beta_0 = \sum_{n=1}^{N} f_n^2(\underline{\phi}) , \quad \beta_1 = \sum_{n=1}^{N} f_n(\underline{\phi}) G_n(\underline{\phi}) ,$$

$$\beta_2 = \sum_{n=1}^{N} \left(f_n(\underline{\phi}) H_n(\underline{\phi}) + G_n(\underline{\phi}) G_n(\underline{\phi})^T\right) . \tag{58}$$

Substitution in equation (54) yields

$$E(\phi) \cong [\mathbf{d}^T \mathbf{d} - 2 \alpha_0 + \beta_0] + 2 [\beta_1 - \alpha_1]^T \Delta + \Delta^T [\beta_2 - \alpha_2] \Delta. \tag{59}$$

Since $\phi = \underline{\phi}$ is the location of a minimum of $E(\phi)$, then

$$\beta_1 - \alpha_1 = \sum_{n=1}^{N} [f_n(\underline{\phi}) - d_n] G_n(\underline{\phi}) = 0 \quad (P\times 1 \text{ vector}) . \tag{60}$$

Also,

$$\beta_2 - \alpha_2 = \sum_{n=1}^{N} \left([f_n(\underline{\phi}) - d_n] H_n(\underline{\phi}) + G_n(\underline{\phi}) G_n(\underline{\phi})^T\right)$$

$$\cong \sum_{n=1}^{N} G_n(\underline{\phi}) G_n(\underline{\phi})^T , \tag{61}$$

as will be argued shortly. It follows that

$$E(\phi) \cong [d^T d - 2\alpha_0 + \beta_0] + \Delta^T \sum_{n=1}^{N} G_n(\underline{\phi})\, G_n(\underline{\phi})^T \Delta$$

$$= [d - f(\underline{\phi})]^T [d - f(\underline{\phi})] + \Delta^T \sum_{n=1}^{N} G_n(\underline{\phi})\, G_n(\underline{\phi})^T \Delta \,. \quad (62)$$

Comparison of equation (62) with equation (49) yields

$$\frac{1}{2} H_E(\underline{\phi}) \cong \sum_{n=1}^{N} G_n(\underline{\phi})\, G_n(\underline{\phi})^T \equiv Q = Q^T \,. \quad (63)$$

The rank of PxP matrix Q is P because the number N of data points **d** is generally much larger than the total number P of unknown parameters ϕ. This result affords a method of computing the Hessian matrix $H_E(\phi)$ of total error $E(\phi)$ at its minimum location $\phi = \underline{\phi}$. It involves only the gradient vectors $\{G_n(\phi)\}$, n=1:N, of the component basis functions $\{f_n(\phi)\}$ (see equation (51)). This result is similar to that in reference 1 (pages 682 - 683).

To apply these results to a practical case, a particular form is presumed for the received data **d**, namely,

$$d_n = f_n(\phi_o) + w_n \quad \text{for } n=1:N \,, \quad (64)$$

where ϕ_o is the true value of the parameter vector and $\{w_n\}$ is additive noise. Then, equation (60) yields

$$\sum_{n=1}^{N} [f_n(\phi_o) - f_n(\underline{\phi}) + w_n]\, G_n(\underline{\phi}) = 0 \quad (P\times 1) \,. \quad (65)$$

However, for high signal-to-noise ratio in equation (64), $\underline{\phi}$ is near ϕ_o, allowing for approximation

$$f_n(\phi_o) \cong f_n(\underline{\phi}) + G_n(\underline{\phi})^T (\phi_o - \underline{\phi}) \tag{66}$$

and, therefore, from equation (65),

$$\sum_{n=1}^{N} G_n(\underline{\phi}) \, [G_n(\underline{\phi})^T (\phi_o - \underline{\phi}) + w_n] \cong 0 \quad (P\times1) \; . \tag{67}$$

That is,

$$Q\,(\phi_o - \underline{\phi}) + \sum_{n=1}^{N} w_n \, G_n(\underline{\phi}) \cong 0 \; , \tag{68}$$

or

$$\underline{\phi} \cong \phi_o + Q^{-1} \sum_{n=1}^{N} w_n \, G_n(\underline{\phi}) \; , \tag{69}$$

where matrix Q is given by equation (63).

The location $\underline{\phi}$ of the minimum of error $E(\phi)$ in equation (48) is a random variable; therefore, matrix Q defined in equation (63) is also random. However, for a large number N of data points and high input signal-to-noise ratio, minimum location $\underline{\phi}$ will not fluctuate much. Then, a reasonable approximation in equation (69) is that the dominant perturbation is caused by the additive noise term $\{w_n\}$, while the quantities Q and $\{G_n(\underline{\phi})\}$ are relatively constant. Under this assumption, the mean-square error matrix of $\underline{\phi}$ is, from equation (69),

$$MSE(\underline{\phi}) = E\{(\underline{\phi} - \phi_o)(\underline{\phi} - \phi_o)^T\} \cong Q^{-1} \sum_{n,m=1}^{N} G_n(\underline{\phi}) \; c_{nm} \; G_m(\underline{\phi})^T \; Q^{-1} , \tag{70}$$

where $c_{nm} = E\{w_n\ w_m\}$, n,m=1:N, are the covariance matrix elements of the additive noise. Since the additive noise covariance matrix $[c_{nm}]$ can be measured apriori, and gradients $\{G_n(\underline{\phi})\}$, n=1:N, as well as matrix Q can be calculated once the solution point $\underline{\phi}$ for minimum $E(\phi)$ is found, the mean-square matrix $MSE(\underline{\phi})$ can be evaluated at solution $\underline{\phi}$ by means of equation (70).

For the special case of white noise $\{w_n\}$, $c_{nm} = \sigma_w^2\ \delta_{nm}$, and it follows that

$$MSE(\underline{\phi}) \cong Q^{-1} \sum_{n=1}^{N} \sigma_w^2\ G_n(\underline{\phi})\ G_n(\underline{\phi})^T\ Q^{-1} = \sigma_w^2\ Q^{-1} \cong 2\ \sigma_w^2\ H_E(\underline{\phi})^{-1}\ , \tag{71}$$

upon use of equation (63). (This result agrees with reference 1, equations (15.5.15) and (15.5.11); however, equation (70) is a more general result for any additive noise covariance matrix.) In particular, the mean-square error of the p-th parameter estimate $\underline{\phi}_p$ is $2\ \sigma_w^2$ times the p-th diagonal of the inverse of the Hessian matrix of total error $E(\phi)$ at the solution point $\underline{\phi}$.

From equation (63), it follows that

$$Q = \sum_{n=1}^{N} G_n(\underline{\phi})\ G_n(\underline{\phi})^T \cong \frac{1}{2}\ H_E(\underline{\phi})\ . \tag{72}$$

The first form requires calculation of N Px1 gradient vectors $\{G_n(\phi)\}$ of signal components $\{f_n(\phi)\}$ in equation (51), whereas the last form requires calculation of the PxP Hessian matrix of scalar error $E(\phi)$ in equation (48). The end result in equation

(72) is a useful approximation only for white noise $\{w_n\}$ and high signal-to-noise ratio in form (64).

In equation (61), a term was dropped, namely,

$$\sum_{n=1}^{N} [f_n(\underline{\phi}) - d_n] \, H_n(\underline{\phi}) \; . \tag{73}$$

For received data model (64), this term becomes

$$\sum_{n=1}^{N} [f_n(\underline{\phi}) - f_n(\phi_o) - w_n] \, H_n(\underline{\phi}) \; . \tag{74}$$

For high signal-to-noise ratio, $\underline{\phi}$ will be close to ϕ_o, and this term is essentially a sum of random noise × signal terms. (This is true for the general case in equation (73) as well.) The remaining term in equation (61), namely,

$$\sum_{n=1}^{N} G_n(\underline{\phi}) \, G_n(\underline{\phi})^T \; , \tag{75}$$

is a sum of signal × signal terms. Therefore, for high signal-to-noise ratio, it is expected that this latter term will dominate and that the approximation in equation (61) is valid (see also reference 1, page 683, especially equation (15.5.11), regarding this topic).

CHANNEL AND WAVEFORM CHARACTERIZATION

TIME-VARYING IMPULSE RESPONSE

Consider a transmitted signal consisting of a unit-impulse transmitted at time t_1, namely $\delta(t - t_1)$, and let the received signal at the channel output be $A\ \delta(t - t_2)$, where time

$$t_2 = f(t_1) \geq t_1 . \tag{76}$$

The inverse function to f is $\tilde{f}$, where

$$t_1 = \tilde{f}(t_2) \leq t_2 . \tag{77}$$

Then the time-varying impulse response of this channel at time t, due to a unit-impulse excitation at time t_1, is

$$h(t;t_1) = A\ \delta(t - f(t_1)) . \tag{78}$$

More generally, for an arbitrary waveform u(t) transmitted through this channel, the received waveform is

$$r(t) = \int dt_1\ u(t_1)\ h(t;t_1) = A \int dt_1\ u(t_1)\ \delta(t - f(t_1))$$

$$= A \int dt_2\ \tilde{f}'(t_2)\ u(\tilde{f}(t_2))\ \delta(t - t_2) = A\ \tilde{f}'(t)\ u(\tilde{f}(t)) , \tag{79}$$

upon using the change of variable $t_1 = \tilde{f}(t_2)$. Usually, to a very good approximation, $\tilde{f}'(t)$ is constant within the observation interval, thereby yielding received waveform

$$r(t) \cong A\ u(\tilde{f}(t)) , \tag{80}$$

where scaling A absorbs this constant factor. From equation

(77), the inverse function satisfies the rule $\tilde{f}(t) \leq t$ for all t. Thus, to find the received waveform, it is necessary to determine the time delay function f( ) in equation (76) and then find its inverse $\tilde{f}($ $)$ according to equation (77).

MOVING SOURCE AND MOVING RECEIVING ELEMENTS

Let the source be at location x(t), y(t), z(t) at time t, while receiving element m of an array is at $x_m(t)$, $y_m(t)$, $z_m(t)$. Then, for a direct path, straight-line transmission of an impulse at time t_1 and speed of propagation c, it follows that

$$c^2 (t_2 - t_1)^2 = [x(t_1) - x_m(t_2)]^2 + [y(t_1) - y_m(t_2)]^2 + [z(t_1) - z_m(t_2)]^2 . \tag{81}$$

The left side of this equation is the square of the distance traveled by an emitted impulse between transmission at time t_1 and reception at time t_2. The right side is the square of the distance between the source location at emission time t_1 and the m-th receiving element location at reception time t_2.

To solve equation (81) for t_1 in terms of t_2, it is necessary to specify the forms of source location functions x(t), y(t), z(t), as well as the location functions of the m-th receiving element. For a constant-depth source with constant velocity, it follows that

$$x(t) = x + \dot{x}\, t \ , \quad y(t) = y + \dot{y}\, t \ , \quad z(t) = z \ , \tag{82}$$

where all five parameters are unknown. For a receiving array moving in the direction of the x-axis, at constant depth and velocity, it follows for the m-th element that

$$x_m(t) = x_m + v\, t \ , \quad y_m(t) = y_m \ , \quad z_m(t) = z_m \ , \tag{83}$$

where all these parameters are presumed known at the receiver. Substitution of equations (82) and (83) in equation (81) yields

$$\begin{aligned} c^2\,(t_2 - t_1)^2 &= (x + \dot{x}\, t_1 - x_m - v\, t_2)^2 \\ &\quad + (y + \dot{y}\, t_1 - y_m)^2 + (z - z_m)^2 \ . \end{aligned} \tag{84}$$

That is,

$$t_1^2\, \alpha - 2\, t_1\, \beta_m(t_2) + \gamma_m(t_2) = 0 \ , \tag{85}$$

where

$$\begin{aligned} \alpha &= c^2 - \dot{x}^2 - \dot{y}^2 \ , \\ \beta_m(t) &= c^2\, t + \dot{x}\,(x - x_m - v\, t) + \dot{y}\,(y - y_m) \ , \\ \gamma_m(t) &= c^2\, t^2 - q_m(t) \ , \\ q_m(t) &= (x - x_m - v\, t)^2 + (y - y_m)^2 + (z - z_m)^2 \ . \end{aligned} \tag{86}$$

Since equation (85) is quadratic in t_1, it has explicit solution

$$t_{1m} = \frac{\beta_m(t_2) - \left(\beta_m^2(t_2) - \alpha\, \gamma_m(t_2)\right)^{\frac{1}{2}}}{\alpha} \ , \tag{87}$$

where the negative square root must be taken to ensure that

$t_{1m} \le t_2$. The dependence of time t_1 on element number m is made explicit here. Therefore, from equation (77), the m-th inverse function is

$$\tilde{f}_m(t) = \frac{\beta_m(t) - \left(\beta_m^2(t) - \alpha\,\gamma_m(t)\right)^{\frac{1}{2}}}{\alpha} \quad \text{for all } m \,. \tag{88}$$

At this point, it is useful to generalize and also include a surface bounce path as well as a bottom bounce path. Thus, $\tilde{f}_m(t)$ will be denoted by $d_m(t)$ for a direct path, by $s_m(t)$ for a surface path, and by $b_m(t)$ for a bottom path. The function $d_m(t)$ uses argument $z - z_m$ in the last term of $q_m(t)$, as already indicated in the bottom line of equation (86), whereas $s_m(t)$ uses argument $z + z_m$ instead, and $b_m(t)$ uses argument $z + z_m - 2\,d$, where d is the water depth. Also, define

$$X_m(t) = x - x_m - v\,t \,, \qquad Y_m = y - y_m \,,$$

$$Z_m = \left\{\begin{array}{ll} z - z_m & \text{for direct path} \\ z + z_m & \text{for surface path} \\ z + z_m - 2\,d & \text{for bottom path} \end{array}\right\} . \tag{89}$$

Then, by canceling c^4 terms to maintain significance, the discriminant in equation (88) becomes

$$\begin{aligned} D_m(t) \equiv \beta_m^2(t) - \alpha\,\gamma_m(t) &= c^2 \left(X_m(t) + \dot{x}\,t\right)^2 + c^2 \left(Y_m + \dot{y}\,t\right)^2 \\ &\quad - \left(\dot{y}\,X_m(t) - \dot{x}\,Y_m\right)^2 + \left(c^2 - \dot{x}^2 - \dot{y}^2\right) Z_m^2 \,, \end{aligned} \tag{90}$$

while, from equations (86) and (89),

$$\beta_m(t) = c^2\, t + \dot{x}\, X_m(t) + \dot{y}\, Y_m \,, \tag{91}$$

which is independent of Z_m. By combining these results, equation (88) can be expressed in its final form as

$$\tilde{f}_m(t) = \frac{c^2\, t + \dot{x}\, X_m(t) + \dot{y}\, Y_m - D_m(t)^{1/2}}{c^2 - \dot{x}^2 - \dot{y}^2} \,, \tag{92}$$

with Z_m and $D_m(t)$ given by equations (89) and (90), respectively, for the case of interest. In these results, there are no assumptions about $\dot{x}$, $\dot{y}$, or v being small relative to sound speed c. However, to order 1/c,

$$\tilde{f}_m(t) \cong t - \frac{1}{c} \left[\left(X_m(t) + \dot{x}\, t\right)^2 + \left(Y_m + \dot{y}\, t\right)^2 + Z_m^2 \right]^{1/2} , \tag{93}$$

if an approximation is desired.

As a special case, a stationary source has $\dot{x} = \dot{y} = 0$, giving exactly

$$\alpha = c^2 \,, \quad \beta_m(t) = c^2\, t \,, \quad \gamma_m(t) = c^2\, t^2 - q_m(t) \,,$$

$$D_m(t) = c^2\, q_m(t) \,, \quad \tilde{f}_m(t) = t - \frac{1}{c}\, q_m(t)^{1/2} \,. \tag{94}$$

The quantity $q_m(t)$ follows from equation (86) as

$$q_m(t) = (x - x_m - v\, t)^2 + (y - y_m)^2 + Z_m^2 \,, \tag{95}$$

where Z_m is given generally by equation (89).

TRANSMITTED AND RECEIVED WAVEFORMS

The transmitted waveform is taken as

$$u(t) = e(t - T)\ g \cos(\omega\ t + p)$$

$$= e(t - T)\ [I \cos(\omega\ t) + Q \sin(\omega\ t)]\ , \tag{96}$$

where I, Q, T, ω are unknown to the receiver. It is presumed that envelope e(t) of u(t) is known, at least approximately. The envelope duration is L seconds. Also, without loss of generality, envelope e(t) starts at zero at time t = 0, and max{e(t)} = 1.

For a direct, surface, and bottom path, the model of the received signal waveform at the m-th receiving element, m=1:M, is taken as

$$\begin{aligned} r_m(t) = {} & e\Big(d_m(t) - T\Big)\ \Big[I_d \cos\Big(\omega\ d_m(t)\Big) + Q_d \sin\Big(\omega\ d_m(t)\Big)\Big] \\ & + e\Big(s_m(t) - T\Big)\ \Big[I_s \cos\Big(\omega\ s_m(t)\Big) + Q_s \sin\Big(\omega\ s_m(t)\Big)\Big] \\ & + e\Big(b_m(t) - T\Big)\ \Big[I_b \cos\Big(\omega\ b_m(t)\Big) + Q_b \sin\Big(\omega\ b_m(t)\Big)\Big] \end{aligned} \tag{97}$$

where I_d, Q_d, I_s, Q_s, I_b, Q_b, T, ω are unknown, in addition to the source parameters x, y, z, $\dot{x}$, $\dot{y}$ that appear in $\{d_m(t)\}$, $\{s_m(t)\}$, and $\{b_m(t)\}$. These latter three sets of functions are available from equations (92) and (90), when combined with the corresponding $\{Z_m\}$ parts of equation (89). The receiving array parameters $\{x_m\}$, $\{y_m\}$, $\{z_m\}$ for m=1:M and receiver speed v are assumed known, as is sound speed c (see equation (83)).

Let $t = 0$ correspond to the beginning of the first signal arrival (direct path) at element number 1 of the receiving array. Then, $e(0) = 0$ yields $d_1(0) - T = 0$, or $T = d_1(0) < 0$, giving

$$
\begin{aligned}
r_m(t) = {} & e\left(d_m(t) - d_1(0)\right) \left[I_d \cos\left(\omega\, d_m(t)\right) + Q_d \sin\left(\omega\, d_m(t)\right)\right] \\
& + e\left(s_m(t) - d_1(0)\right) \left[I_s \cos\left(\omega\, s_m(t)\right) + Q_s \sin\left(\omega\, s_m(t)\right)\right] \\
& + e\left(b_m(t) - d_1(0)\right) \left[I_b \cos\left(\omega\, b_m(t)\right) + Q_b \sin\left(\omega\, b_m(t)\right)\right] \qquad (98)
\end{aligned}
$$

for $m=1:M$. Unknowns I_d, Q_d, I_s, Q_s, I_b, Q_b appear linearly, while unknowns x, y, z, $\dot{x}$, $\dot{y}$, ω appear nonlinearly in modeled received signals $\{r_m(t)\}$ through the $\{d_m(t)\}$, $\{s_m(t)\}$, and $\{b_m(t)\}$ functions.

The modeled received data are sampled at times $t = n\Delta$ for $n=1:N$, giving model data

$$
\begin{aligned}
r_m(n\Delta) = {} & e\left(d_m(n\Delta) - d_1(\Delta)\right) \left[I_d \cos\left(\omega\, d_m(n\Delta)\right) + Q_d \sin\left(\omega\, d_m(n\Delta)\right)\right] \\
& + e\left(s_m(n\Delta) - d_1(\Delta)\right) \left[I_s \cos\left(\omega\, s_m(n\Delta)\right) + Q_s \sin\left(\omega\, s_m(n\Delta)\right)\right] \\
& + e\left(b_m(n\Delta) - d_1(\Delta)\right) \left[I_b \cos\left(\omega\, b_m(n\Delta)\right) + Q_b \sin\left(\omega\, b_m(n\Delta)\right)\right] \qquad (99)
\end{aligned}
$$

for $m=1:M$ and $n=1:N$. This model is to be fit to the actual measured data $\{p_m(n\Delta)\}$ at element number m and time $n\Delta$, by choosing the following quantities:

$$I_d,\ Q_d,\ I_s,\ Q_s,\ I_b,\ Q_b,\ x,\ y,\ z,\ \dot{x},\ \dot{y},\ \omega\ . \tag{100}$$

The first six variables are to be eliminated analytically, as indicated in an earlier section. This elimination reduces the number of search dimensions from 12 to 6, a very worthwhile procedure. A numerical search on the remaining six variables in equation (100) is then required.

When estimates of x, y, z, $\dot{x}$, $\dot{y}$ become available after data processing, they constitute position and velocity estimates of the source at the time $t = 0$ (see equation (82)) when the direct-path signal first arrived at element number 1. This analysis presumes that the source kept a steady course from the initiation of its transmission at time T (see equation (96)) until time $t = 0$. The quantity $T = d_1(0)$ is negative and is available from equations (89) through (92) in the form

$$T = d_1(0) = \frac{\dot{x}\,(x - x_1) + \dot{y}\,(y - y_1) - D_1(0)^{\frac{1}{2}}}{c^2 - \dot{x}^2 - \dot{y}^2}\ , \tag{101}$$

with

$$\begin{aligned} D_1(0) = {} & c^2\,(x - x_1)^2 + c^2\,(y - y_1)^2 \\ & - [\dot{y}\,(x - x_1) - \dot{x}\,(y - y_1)]^2 + (c^2 - \dot{x}^2 - \dot{y}^2)\,(z - z_1)^2\ . \end{aligned} \tag{102}$$

The array positions $\{x_m\}$, $\{y_m\}$, $\{z_m\}$ are the locations of the receiving elements at time $t = 0$ (see equation (83)).

The time estimate T is available once x, y, z, $\dot{x}$, $\dot{y}$ are available. The estimated source position at time T (<0) is

$$P(T) \equiv [x + \dot{x}\,T \,, \quad y + \dot{y}\,T \,, \quad z] \,. \tag{103}$$

This source position estimate at time T, namely P(T), is the only reliable description of the actual source path because the actual source path could have deviated from the straight-line assumption utilized in equation (82), both before and after the emission time T. Since there was no emission from the source before time T or after time T+L, there is no information about the actual source path at times other than those in the time interval [T,T+L]. Source position estimates at other times are only projections based on the constant heading assumption.

SUMMARY

A method for reducing the dimensionality of the search for a minimum in several parameters has been derived for the case where some of the parameters appear linearly in the model of the observed data. The remaining nonlinearly-appearing parameters in the model must still be searched in multidimensional space. The advantages in execution time achieved by this approach can easily be several orders of magnitude.

Elimination of the linearly-appearing parameters significantly complicates the conditionally-optimum total squared error, resulting in a Hermitian form that must be extremized. For efficient searching in the resultant space, the gradient vector and the Hessian matrix of this Hermitian form must be calculated. Expressions for both of these quantities have been derived.

The Hessian matrix was found to contain a destabilizing term, just as there is for the standard least-squares approach on all the parameters. The form of this term, however, is different from that encountered in the simpler standard approach. Nevertheless, no second-order derivatives of the basis functions need be computed, thereby significantly reducing the amount of computer effort required to evaluate the Hessian matrix.

REFERENCES

1. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, **Numerical Recipes**, Cambridge University Press, New York, 1988.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Naval Sea Systems Command (PEO IWS/ASTO (R. Dostie, R. Zarnich); PMS 415 (S. Von Christierson))	3
Defense Technical Information Center	2
Center for Naval Analyses	1